Am J Hum Genet 37:362–372, 1985

Paternal-Age and Birth-Order Effect on the Human Secondary Sex Ratio

AVIMA RUDER[1]

SUMMARY

Because of conflicting results in previous analyses of possible maternal and paternal effects on the variation in sex ratio at birth, records of United States live births in 1975 were sorted by offspring sex, live birth order (based on maternal parity), parental races, and, unlike prior studies, ungrouped parental ages. Linear regression and logistic analysis showed significant effects of birth order and paternal age on sex ratio in the white race data (1.67 million births; 10,219 different combinations of independent variables). Contrary to previous reported results, the paternal-age effect cannot be ascribed wholly to the high correlation between paternal age and birth order as maternal age, even more highly correlated with birth order, does not account for a significant additional reduction in sex-ratio variation over that accounted for by birth order alone.

INTRODUCTION

The human secondary sex ratio (male births/all births) is a subject of scientific interest due to the influence of natural selection and effect of radiation on the sex ratio, the use of sex-ratio changes to estimate mutation rates, and the demographic importance of sex-ratio studies [1].

Since even interracial variation in sex ratio fluctuates only between 0.485 and 0.530, less than 5%, only with a very large number of birth statistics can most variation safely be attributed to effects of the factors being studied rather than

Received June 19, 1984; revised October 16, 1984.

This study was done in partial fulfillment of the requirements for the Ph.D. degree of the University of Oregon, Eugene. This investigation was supported by grants 5-T32-GM7413-05 from the U.S. Public Health Service, S07-RR07080 from the Biomedical Research Support Grant Program, National Institutes of Health, and 5-R01-GM18678-11 from the National Institutes of Health, and by University of Oregon Computing Center research funds.

[1] Department of Clinical Epidemiology, Chaim Sheba Medical Center, affiliated with Tel Aviv University Sackler Medical School, Tel Hashomer 52621, Israel.

to small-size sample fluctuation. Many reports on sex ratio have dealt with sample sizes inadequate for statistical purposes.

Nearly all the large-scale statistical studies of human births have found significant correlation of sex ratio with birth order. Otherwise, they differ in data sources, approaches, and conclusions (table 1). Relationships have been shown—and disputed—for maternal age, paternal age, and birth order with the variation in sex ratio. Analysis is confounded by the high correlations among these three factors.

In an early large-scale analysis, Novitski [2] found an association between increasing paternal age and decreasing sex ratio. Showing that declining sex ratio correlates more closely with advancing paternal than advancing maternal age, Novitski and Sandler [3] proposed that a biological basis for the changing sex ratio might be a shift in the relative frequency of X- and Y-bearing sperm.

Novitski and Kimball [4] tested parental ages and birth order as independent variables. Their best-fitting quadratic regression models included birth order, paternal age, and powers or combinations of these two factors as variables. The Novitski and Kimball study was limited because their data source, the U.S. Bureau of Vital Statistics, had grouped parental ages in 5-year classes, with oldest and youngest ages truncated. Reanalyzing the same data, Teitelbaum et al. [6] compared mean birth order and parental ages of females and males with Mantel-Haenzel contingency tables [10], finding only a birth order effect. Contingency tables compare means of two or more sample subpopulations, assuming that the data vary about the mean in a random fashion. If monotonic trends exist within groups, a contingency table analysis may not be adequate [11].

Erickson [7] found a significant partial regression coefficient only for birth order, but he pointed out that since birth order accounted for only 10% of the sex-ratio variation in his data, other factors must play a part. Multivariate analyses of white, singleton, live births by Garfinkel and Selvin [8] indicated birth order plus parental ages did not account for significantly more variation than explained by birth order alone.

Data in all these U.S. studies were truncated or grouped, especially for older and younger parental ages, usually before they were made available to the analysts. Age cutoffs or groupings can severely bias a multivariate analysis [12].

METHODOLOGY

The United States National Center for Health Statistics (NCHS) natality data tapes for over 2 million live births in 1975 include demographic details (personal identification removed) from 100% of the birth certificates filed in 33 states and a 50% sample from the other states [13, 14].

We developed a series of FORTRAN computer programs to abridge, sort, and analyze this large number of records on the University of Oregon Computing Center IBM-360. While classifying records as: multiple births, births to older (≥ 45) or younger (< 15) parents, missing age or race data, intraracial, and interracial, our program TRANSFER eliminated extraneous details to include only mother's residence and education; parents' races and ages; interval since and outcome of last pregnancy; gestation length; total and live birth order; birthplace, birthdate, child's sex, and birthweight; and plurality.

TABLE 1

LARGE-SCALE STUDIES OF SEX RATIO

Source	Database	Data cells	Method	Best model	Limits*
Novitski [2]	U.S. 1947–48	90	Linear regression	Paternal age	Birth orders, races combined
Novitski and Sandler [3]	U.S. whites, 1947–52	90	Weighted linear regression	Birth order, paternal age	Ages grouped
Novitski and Kimball [4]	U.S. 1955	175	Quadratic regression	Birth order, paternal age	Ages grouped, races combined
Pollard [5]	Australia 1902–65	36	Regression, chi-square	Parental age	Ages grouped
Teitelbaum et al. [6]	U.S. 1955	385	Chi square	Birth order	Ages grouped, races combined
Erickson [7]	U.S. 1969–71	336	Chi-square, linear regression	Birth order	Ages grouped
Garfinkel and Selvin [8]	N.Y. whites, 1955–67	7,104	Regression models	Birth order	Ages grouped
Rostron and James [9]	Scotland 1961–72	770	ANOVA	Birth order, maternal age	Ages grouped
Imaizumi and Murata [1]	Japan 1975–76	378	Weighted least squares	None	Ages grouped
Present study	U.S. 1975	25,381	Logistic, linear regression	Paternal age, birth order	...

* Most limitations inherent in government data used.

Another program, SHUFFLE, sorted births into cells by parental ages and child's birth order and saved cells with one or more births. Each contained a unique combination of parental ages and race(s), birth order, and numbers of male and female births. No sorting or analysis has yet been done for the other parameters in the birth records.

Our logistic analysis program, LOGIT, fits models to the data by an algorithm proposed by Berkson [15, 16] and by a less time-efficient but reiterative algorithm. The null hypothesis for logistic analysis follows the binomial distribution. If no independent factors have an effect, the probability of exactly m males in a sample of n births would be $(n!/(m!(n - m)!))P^m(1 - P)^{n-m}$. P is the sex ratio in the population that the sample comes from [17].

Evaluating independent factors, both algorithms sought an intercept (a) and coefficients ($b_1, b_2, \ldots, b_k$) that helped to reduce variation between the observed and expected (estimated) sex ratios. For the ith different combination of the independent factors, the estimated sex ratio is

$$y_i = \frac{\exp(a + b_1x_{1i} + b_2x_{2i} + \ldots + b_kx_{ki})}{1 + \exp(a + b_1x_{1i} + b_2x_{2i} + \ldots + b_kx_{ki})} .$$

The variable employed for the jth factor, x_{ji}, = ith value − mean value.

Other statistical techniques used include: linear regression with the program BMDP1R [18]; chi-square tests; Fisher's test for samples with large numbers of degrees of freedom (n), $P(\chi^2 \geq n) = 0.5$, with deviations distributed on N(0,1) for $z = \sqrt{2\chi^2} - \sqrt{(2n - 1)}$ [19]; t-tests for correlations, $T = [r\sqrt{(n - 2)}]/[\sqrt{(1 - r^2)}]$, with r the correlation coefficient and n the sample size, T having a t distribution with $n - 2$ df; and t-tests of the significance of a coefficient: T = [(coefficient) i/(standard error of i)].

RESULTS

The 11 sorted categories contained records of 2,232,401 births. Over 11% (254,015), missing parental age(s) or race(s), were eliminated from this analysis, as were multiple and interracial births and the Guamian sample, which was too small for meaningful analysis. Numbers of births and data cells, sex ratios, confidence limits, and expected chi-squares are in table 2. Each data cell (one or more observed births with the same unique combination of age and birth-order values) was expected to contribute 1.0 to the total chi-square. Deviation from the expected was calculated with Fisher's test (METHODOLOGY).

Sex ratios found for white and black births were consistent with those reported in earlier U.S. studies. Sex ratios for other races (categories are those used by NCHS [14] to code the data) have not been reported previously for U.S. births. Sex ratios for Chinese, Hawaiian, "Oriental," and Filipino births approximated sex ratios for Asians in Asia reported in the literature. The low American Indian and Japanese sex ratios may have been artifacts of small sample size.

F-values of linear regression analyses by BMDP1R [18] on all births, collectively and by race, are in table 3. The null hypothesis that no independent variable(s) affects sex-ratio variation could be rejected for the white-race data and, because white births were 87% of the total, for the combined data.

Correlations between the three factors being considered as possible independent variables are in table 4. For every correlation, $P \leq .005$, evaluated with t-tests (METHODOLOGY), that such an association could occur by chance.

In logistic analyses, all three factors contributed significantly to im-

TABLE 2

SEX RATIOS, BY RACE, FOR INTRARACIAL BIRTHS

Race	Sample size	No. data cells*	Observed chi-square	Sex ratio	99% confidence interval
White	1,665,457	10,219	10,087	.5149	(.5140, .5158)
Black	217,955	7,786	7,731	.5099	(.5074, .5124)
American Indian	7,416	2,093	2,120	.5011	(.4876, .5146)
Chinese	3,611	1,086	1,054	.5164	(.4970, .5358)
Japanese	1,952	640	642	.5036	(.4772, .5300)
Hawaiian	705	446	449	.5163	(.4725, .5602)
"Oriental"	7,425	1,697	1,674	.5263	(.5128, .5398)
Filipino	3,684	1,341	1,346	.5288	(.5096, .5479)
Guamian	79	73	69	.4304	(.3006, .5602)
Combined	1,908,284	...	...	.5143	(.5135, .5151)

* Each cell represents a unique combination of parental ages and birth order. The no. cells is the expected chi-square. In no case did the observed chi-square differ significantly.

TABLE 3

F VALUES OF LINEAR REGRESSIONS FOR INTRARACIAL SEX RATIOS

	One-factor models			Two-factor models			Three-factor model
Race	Birth order	Paternal age	Maternal age	Parental ages	Birth order and paternal age	Birth order and maternal age	Birth order and paternal ages
White	5.14*	4.44*	2.95	2.23	3.42*	2.87	2.31
Black	0.03	0.28	0.18	1.04	0.25	0.09	0.69
Indian	1.23	2.27	1.28	1.13	1.17	0.74	0.78
Chinese	2.49	3.28	0.13	2.03	2.23	1.25	1.87
Japanese	0.03	1.27	3.38	1.73	0.80	2.22	1.50
Hawaiian	0.07	1.41	0.67	0.71	0.83	0.39	0.55
"Oriental"	1.17	0.09	0.47	0.25	0.80	1.23	0.82
Filipino	0.64	2.49	2.09	1.52	1.35	1.09	1.03
Combined	6.44*	5.10*	1.72	2.80	4.13*	3.24*	3.22*

NOTE: Critical F values for large samples (1,000+ cases) at 95% significance level: 3.85 (1 variable), 3.00 (2), 2.61 (3).
* Significant at $P \leq .05$.

TABLE 4

CORRELATIONS BETWEEN FACTORS CONSIDERED FOR POSSIBLE EFFECT ON SEX RATIO

Race of parents	Parental ages	Maternal age/birth order	Paternal age/birth order
White	.7787	.4544	.4008
Black	.7794	.5300	.4622
Indian	.7480	.6995	.5941
Chinese	.5935	.2961	.2977
Japanese	.7243	.4135	.3274
Hawaiian	.7774	.6465	.5769
"Oriental"	.6479	.3493	.3102
Filipino	.5102	.3685	.2294
Combined	.7775	.4564	.4047

NOTE: All correlations were significant at $P \leq .005$.

provement of goodness of fit of the white-race data (table 5). Coefficients for the statistically significant logistic models were evaluated with *t*-tests (METHODOLOGY); results are in table 6. An estimated sex ratio for any combination of independent values can be obtained by substituting the variable(s) and the coefficient(s) into the logistic model (METHODOLOGY).

Mean birth orders of males and females and mean ages of parents of daughters and sons were compared with chi-square tests, the null hypotheses being that no differences exist. The results (table 7) were significant for all three factors.

DISCUSSION

The white-race linear regression and logistic analysis results show that both paternal age and birth order (or biological factors closely correlated with them) play a significant part in determining secondary sex ratio.

TABLE 5

LOGISTIC ANALYSIS FOR INTRARACIAL SEX RATIOS

Race	Birth order	Paternal age	Maternal age	Parental ages	Birth order + Paternal age	Birth order + maternal age	Birth order + Parental ages
White	7.7*	6.5†	4.7†	7.2†	10.1*	3.5	9.8†
Black	0	0.2	0	0	0	0	0
Indian	0.7	0.8	0.6	1.3	1.2	0.9	0.9
Chinese	2.9	1.9	0	1.7	3.6	4.1	4.7
Japanese	0	0.7	2.1	2.1	0.7	1.8	1.8
Hawaiian	0.1	1.1	0.6	1.0	1.1	0.4	0.5
"Oriental"	0.1	0.3	0.4	0.6	0.3	0.3	0.5
Filipino	0.7	2.1	1.4	2.5	2.1	1.1	2.4

NOTE: Chi-square improvement (from unimproved [starting] chi-squares [table 2]) due to independent variables.
* 99% significance level.
† 95% significance level.

TABLE 6

COEFFICIENTS FOR WHITE-RACE LOGISTIC MODELS

Variable	Coefficient (b)	Standard error (Sb)	T-value (b/Sb)
One factor:			
Birth order	−0.00274	0.00105	2.62*
Paternal age	−0.00082	0.00026	3.12*
Maternal age	−0.00065	0.00031	2.09†
Two factor:			
Birth order + ...	−0.00146	0.00105	1.40
Paternal age	−0.00048	0.00026	1.83†
Paternal age + ...	−0.00048	0.00026	1.83†
Maternal age	−0.00019	0.00031	0.62
Three factor:			
Birth order + ...	−0.00274	0.00105	2.62*
Paternal age + ...	−0.00052	0.00026	1.96†
Maternal age	0.00023	0.00031	0.73

NOTE: In all cases, intercept (constant) = 0.05916 = ln (p/1 − p)), with p = 0.5149, the white race sample mean sex ratio.
* 99% significance level (10, 217 df).
† 95% significance level.

In the white-race logistic analysis, birth order plus paternal age or paternal age plus maternal age significantly improved goodness of fit; birth order plus maternal age did not. Therefore, the improvement observed when maternal age but not birth order is one of the independent variable(s) can be attributed to the strong correlation of maternal age and birth order.

Could all parental-age effects observed be due only to age correlations with birth order? If so, we should observe a maternal-age "effect" greater than the

TABLE 7

MEAN AGES OF PARENTS OF AND BIRTH ORDERS OF DAUGHTERS AND SONS

Race	Offspring sex	Sample size	Birth order	Maternal age	Paternal age
White	Daughters	1,665,457	2.123	24.95	27.56
	Sons		2.117*	24.93*	27.54*
Black	Daughters	217,955	2.557	24.05	27.14
	Sons		2.558>	24.06>	27.12
Indian	Daughters	7,416	2.937	24.64	27.74
	Sons		2.883*	24.48*	27.49*
Chinese	Daughters	3,611	1.944	28.29	33.03
	Sons		1.873*	28.24	32.68*
Japanese	Daughters	1,952	1.815	28.90	31.86
	Sons		1.823>	28.56*	31.60†
Hawaiian	Daughters	705	2.575	24.28	27.38
	Sons		2.541	23.96	26.79*
"Oriental"	Daughters	7,425	2.147	27.77	31.87
	Sons		2.190>†	27.70	31.83
Filipino	Daughters	3,684	2.101	28.93	32.98
	Sons		2.065	28.69*	32.51*

NOTE: > = Mean birth order/maternal age of sons > that of daughters.
* $P \leq .01$.
† $P \leq .05$.

observed paternal-age effect since the correlation between maternal age and birth order was higher than that of paternal age and birth order. However, observed maternal age effect was *less* than observed paternal-age effect. If no paternal-age effect existed, then birth order plus paternal age should not have improved goodness of fit any more than birth order alone or birth order plus maternal age. However, birth order plus paternal age improved goodness of fit as significantly with 2 df as did birth order alone with 1 df. Contrariwise, birth order plus maternal age did not significantly improve goodness of fit (table 5).

More support for including paternal age and excluding maternal age among significant factors in the samples studied is provided by the following:

(1) No maternal-age coefficient larger than 10^{-5} improved goodness of fit, in a logistic model already incorporating a nonzero birth-order coefficient, when the algorithm was changed to seek reduction in variation below the *lowest* chi-square value reached, instead of below the original chi-square value.

(2) *T*-tests of coefficients in all three one-factor white-race logistic models were significant (table 6). However, in the two-factor models, only the paternal-age coefficient was significant; in the three-factor model, the maternal-age coefficient was not significant. Because of greater variance in paternal age than in birth-order values, a paternal-age value was more likely to be farther from its mean and the corresponding paternal-age variable (*i*th value − mean value) larger. Therefore, the weight of the paternal-age component (coefficient times *i*th variable) is comparable to that of birth order.

(3) Chi-square tests comparing the mean birth orders and ages of parents of males and females were significant for all three factors (table 7), but the direction of the difference was consistent only for paternal ages.

Previous linear regression analyses presumed a continuous normally distributed dependent variable (sex ratio). But for any one birth, the true dependent variable is the child's sex, and logistic analysis, which assumes a nominal dependent variable with two mutually exclusive outcomes, such as female or male sex, appears to be more appropriate. In addition, a sigmoidal curve may have greater meaning for biological relationships than does a straight line.

Likelihood tests were performed for all logistic analyses; the likelihoods did not improve significantly with an algorithm designed to reduce chi-square. An algorithm designed to maximize likelihood might produce different results.

The improvement in goodness of fit (table 5) for birth order alone is greater than that for birth order plus maternal age, and the improvement with a three-factor model is less than that for birth order plus paternal age. This is because a birth-order coefficient that improved goodness of fit was held constant while the algorithm sought a nonzero maternal-age coefficient that reduced chi-square variation below the *original* (but not the modified) value. As discussed above, modified algorithms could not find acceptable maternal-age coefficients.

The homogeneity of the white- and black-race data might be a peculiarity of the 1975 sample. However, earlier studies used vital statistics data noting only maternal race and grouped by age classes. Inadvertent inclusion of interracial births could increase data heterogeneity; grouping by age classes may have increased apparent heterogeneity [12]. For example, total χ^2 for the 76 largest cells in our data = 83.7 ($P \leqslant .25$, 76 df). If the data are clustered by 5-year paternal-age classes, the 10 groups have total $\chi^2 = 15.5$ ($P \leqslant .12$, 10 df). A decrease in probability ordinarily is interpreted as an increase in heterogeneity. The discrepancy between our chi-square results (table 7) and those of Teitelbaum et al. [6], who observed no paternal-age effect, might also be due to the censored, grouped, and truncated government data they worked with.

The high correlations (table 4) between the three factors evaluated for their effect on the sex ratio mean that to evaluate individual effects one would need stratified samples. One could control for birth order in a sample of first births large enough so any variation-reducing effects could be observed. Similarly, in a sufficiently large sample of first births to mothers of the same age, any observed paternal-age effect could be attributed, with some confidence, to a paternal factor. It would be of interest to combine several years of birth records to assemble a large sample of first births. In our study there were 682,146 white first births. Since only the entire white-race sample, 2½ times larger, yielded significant results, we did not analyze the first births separately.

No study could evaluate simultaneously even a fraction of the factors that have been proposed, over the years, to affect sex-ratio variation. Of these, one might imagine that for biological reasons maternal age in particular would be important. However, since the present analysis, performed on ungrouped data, detected no maternal-age effect, the true number of biologically significant factors affecting sex ratio may be much smaller than has been proposed.

ACKNOWLEDGMENTS

I am most grateful to Edward Novitski for his advice, support, and encouragement throughout the analysis of the data and the preparation of this report. Data tapes, obtained from the National Technical Information Service, originated at the National Center for Health Statistics, which bears no responsibility for any analyses, interpretations, or conclusions based on these data. BMDP Biomedical Computer Programs were developed at the Health Sciences Computing Facility, UCLA, supported by NIH Special Research Resources Grant RR-3.

REFERENCES

1. Imaizumi Y, Murata M: Secondary sex ratio, paternal age, maternal age and birth order in Japan. *Ann Hum Genet* 42:457–465, 1979
2. Novitski E: The dependence of the secondary sex ratio in humans on the age of the father. *Science* 117:531–533, 1953
3. Novitski E, Sandler L: The relationship between parental age, birth order and the secondary sex ratio in humans. *Am J Hum Genet* 21:123–131, 1956
4. Novitski E, Kimball AW: Birth order, parental ages, and sex of offspring. *Am J Hum Genet* 10:268–275, 1958
5. Pollard GN: Factors influencing the sex ratio at birth in Australia. *J Biosoc Sci* 1:125–144, 1969
6. Teitelbaum MS, Mantel N, Stark CR: Limited dependence of the human sex ratio on birth order and parental ages. *Am J Hum Genet* 23:271–280, 1971
7. Erickson JD: The secondary sex ratio in the United States, 1969–71. *Ann Hum Genet* 40:205–212, 1976
8. Garfinkel J, Selvin S: A multivariate analysis of the relationship between parental age and birth order and the human secondary sex ratio. *J Biosoc Sci* 8:113–121, 1976

9. Rostron J, James WH: Maternal age, parity, social class, and sex ratio. *Ann Hum Genet* 41:205–217, 1977
10. Mantel N, Haenszel W: Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 22:719–748, 1959
11. Cox DR: *The Analysis of Binary Data*. London, Methuen, 1970, pp 1–42
12. Caughley G: Offspring sex ratio and age of parents. *J Reprod Fertil* 25:145–147, 1971
13. National Center for Health Statistics: *Standardized MicroData Tape Transcripts*. Hyattsville, Md., NCHS, 1978, pp 9–11
14. National Center for Health Statistics: *Vital Statistics Natality Data, Detail, 1975*. Hyattsville, Md., NCHS, pp 1–27
15. Berkson J: Application of the logistic function to bioassay. *Am Stat Assn J* 39:357–365, 1944
16. Berkson J: Minimum χ^2 and maximum likelihood solution in terms of a linear transform, with particular reference to bioassay. *Am Stat Assn J* 44:273–278, 1949
17. Finney DJ: *Probit Analysis*. Cambridge, England, University Press, 1947, p 15
18. Dixon WJ, Brown MB, eds: *BMDP-79 Biomedical Computer Programs P-Series*. Berkeley, Univ. of California Press, 1979, pp 380–398
19. Fisher RA, Yates F: *Statistical Tables for Biological, Agricultural, and Medical Research*, 6th ed. New York, Hafner, 1964, p 45